INSTITUTE OF PHYSICS PUBLISHING PHYSICS IN MEDICINE AND BIOLOGY

Phys. Med. Biol. **47** (2002) 1–25 PII: S0031-9155(02)35108-X

Parsimonious basis selection in exponential spectral analysis

Jonathan S Maltz

Department of Nuclear Medicine and Functional Imaging, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Mailstop 55-121, University of California, Berkeley, CA 94720, USA

Received 21 March 2002
Published DD MMM 2002
Online at stacks.iop.org/PMB/47/1

Abstract

Sums of decaying real exponentials (SDREs) are frequently used in models of time-varying processes. First-order compartmental models are widely employed to describe mass transit in chemical and biological systems. In these models the evolution of compartment concentration versus time is represented as the convolution of an input function with an SDRE. In exponential spectral analysis (ESA) the nonlinear problem of estimating the SDRE rate constants is replaced by the linear estimation of the coefficients of a preselected set of exponential basis functions (EBFs). This work addresses the problem of selecting the number of EBFs and the rate constant of each basis element. Basis dimension is established via model selection, in which approximation error and parameter redundancy are the criteria. The latter is estimated via simulation of the fitted model over multiple noise realizations. A constrained Cramér–Rao lower bound is derived for ESA parameters. The resulting parsimonious ESA algorithm (PESA) ameliorates the inherent problem of non-uniqueness in ESA parameters. Consequently, sets of time series may be compared in a statistically meaningfully way in terms of physically or physiologically significant parameters. PESA is applied to compare the retention of two radiotracers in the artificially perfused rabbit heart.

1. Introduction

Many physical and physiological processes are readily described using linear first-order models. The output of such models may be represented as the convolution of a forcing function $i(t)$ with an exponential kernel having a rate constant k_2 and amplitude k_1:

$$\phi(t) = i(t) * k_1 k_2 \, \mathrm{e}^{-k_2 t} \tag{1}$$

where the parameter k_2 appears as an amplitude factor in order to effect the normalization

$$\int_0^\infty k_2 \, \mathrm{e}^{-k_2 t} = 1. \tag{2}$$

0031-9155/02/000001+25$30.00

Often the observed quantity $\phi(t)$ comprises the sum of the responses of several first-order systems:

$$\phi(t) = i(t) * \sum_{\tilde{m}=1}^{\tilde{M}} k_1^{\tilde{m}} k_2^{\tilde{m}} \, \mathrm{e}^{-k_2^{\tilde{m}} t}. \tag{3}$$

For example, the concentration versus time curve $\phi(t)$ might represent the superposed responses of several single compartment models describing the kinetics of a certain chemical or pharmaceutical.

We often wish to fit models of the form (3) to sets of measured data. In general, it is not possible to find a unique set of parameters for this model even in the absence of noise (Braess 1986, p 168). Even when uniqueness conditions are met, the parameters are generally extremely sensitive to noise within the data (Lanczos 1956, Reich 1981). This problem, termed 'parameter redundancy', is intrinsic to weighted sums of real exponentials.

When a model of this form is fitted to noisy data, it is very difficult to ascertain the true robustness of the parameter values at the solution, as bounds on parameter variance, such as the Cramér–Rao lower bound, do not account for the variability introduced by the presence of multiple solutions. Since a primary purpose of quantitative modelling is to facilitate comparison of the results of several experiments or of the responses of several systems, models of the form (3) appear inappropriate if this objective is to be realized. What is needed is a method that describes the process modelled in (3) in terms of the same physically meaningful parameters, yet improves the robustness of these parameters so that statistically valid conclusions may be drawn from the results. This paper describes the development of such a method, and its evaluation through application to a tracer kinetic modelling problem in nuclear medicine. While this technique may find broad application in many areas of the physical and biological sciences, such as in modelling the kinetics of stable isotopes or pharmacological agents, we target our method for application to radiotracer studies.

In the following section we begin by illustrating the limitations of direct approaches to fitting the parameters of (3) using nonlinear regression. We then describe methods of approaching the analysing exponential spectra that employ the Laplace transform (LT) and numerical methods of inverting the LT. Existing exponential spectral analysis (ESA) methods, due to Provencher (Provencher 1976, Provencher and Dovi 1979) and Cunningham and Jones (1993) are described in section 4. Section 5 follows with the presentation of the parsimonious basis exponential spectral analysis algorithm (PESA), which is designed to overcome some of the limitations of existing methods. We then demonstrate PESA through application to synthetic and actual experimental data in section 6.

2. Nonlinear parameter estimation

The most direct approach to compartmental model fitting is to perform nonlinear estimation of the model parameters. A familiar way in which this is achieved is through the use of an optimization algorithm to minimize a least-squares cost function. For example, a fit of the compartmental model $\phi(t_l, \mathbf{k_1}, \mathbf{k_2})$ in (3) to an observed time series $\phi'(t_l)$, leads to the cost function

$$\min_{\mathbf{k_1},\mathbf{k_2}} \sum_{l=1}^{L} (\phi'(t_l) - \phi(t_l, \mathbf{k_1}, \mathbf{k_2}))^2 \qquad l = 1, \ldots, L \tag{4}$$

where L is the length of the time series, and $\mathbf{k_1}$ and $\mathbf{k_2}$ are vectors of the wash-in and wash-out parameters, respectively. Optimization algorithms commonly used to solve problems of this

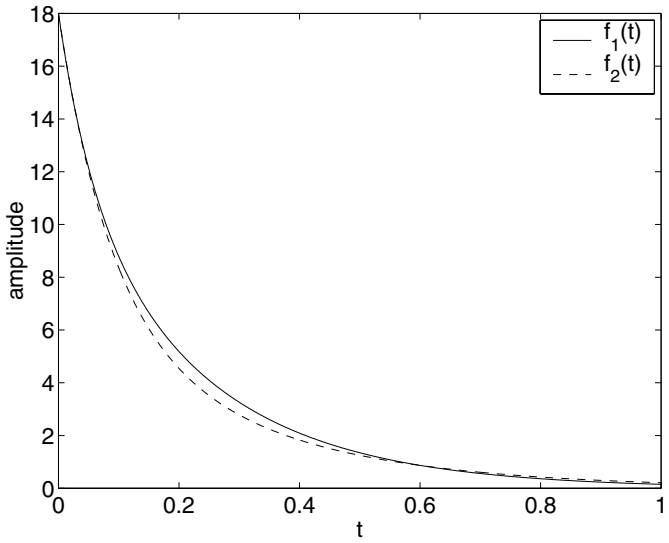


Figure 1. Example of parameter redundancy in e-sums. Here, two functions with widely differing parameters produce curves that are virtually indistinguishable. Shown here are $f_1 = 6\,\mathrm{e}^{-17.5t} + 12\,\mathrm{e}^{-4.375t}$ and $f_2 = 11\,\mathrm{e}^{-11.667t} + 7\,\mathrm{e}^{-3.5t}$.

form include the method of steepest descent, the Newton–Raphson method, conjugate descent methods as well as hybrid methods such as the Levenberg–Marquardt algorithm.

The approach described above has three main disadvantages. The first and the least severe of them is that the number of compartmental responses superposed in a particular measurement is not always known *a priori*.

Secondly, the parameters obtained may be overly sensitive to noise. Reich has shown that errors in e-sum parameter values are ten times the magnitude of the measurement errors in the e-sum when the decay parameters of the components of the e-sum are separated by more than a factor of 5 (Reich 1981). Figure 1 illustrates the phenomenon of parameter redundancy by showing how very similar e-sums (the kernel of the convolution in (3)), may possess very different parameters. Suppose two trials of a single experiment produced these two exponential sums. When examined in terms of the maximum deviation of the two time series from their mean value at any point, an error of 16% is observed. This is far more representative of the similarity of the e-sums than the errors of over 50% obtained through comparison of the rate constants.

Finally, nonlinear estimation of the parameters of (3) is sensitive to the parameter vector used to initialize the estimation algorithm. Let us consider again the function f_1 described in figure 1. Performing a least-squares fit of this function using the Levenberg–Marquardt algorithm starting at $x_0 = \left[k_{10}^1\, k_{10}^2\, k_{20}^1\, k_{20}^2\right] = [0\,0\,0\,0]$ yields a solution of $\hat{x}_a = [8.1887\; 8.1887\; 5.5437\; 5.5437]$. Following the same optimization procedure starting at the point $x_0 = [1\,2\,3\,4]$ yields $\hat{x}_b = [11.9991\; 6.0008\; 4.3748\; 17.4970]$.

We quantify the approximation error using the Tchebycheff norm on a discrete set of sample times

$$\begin{aligned} d_m &\triangleq \max_{t\in T} |\hat{f}_m(t) - f_m(t)| \\ T &\triangleq \{t \mid t = t_l, l = 1, \ldots, L\} \end{aligned} \tag{5}$$

where $\hat{f}_m(t)$ is the approximation to $f_m(t)$ and L is the number of samples.

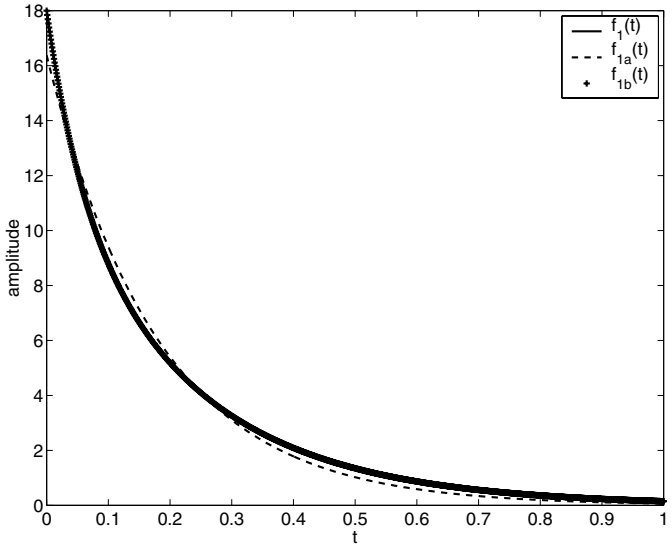


Figure 2. The presence of multiple local optima introduces sensitivity to the parameter estimate used to initialize the parameter optimization algorithm. Here, the parameters of f_1 are estimated starting from two different initial points in parameter space. The function f_{1a} represents a suboptimal fit obtained at a local optimum. The graph of f_{1b} cannot be differentiated from the true curve in this plot.

We find in the fitting of f_1 that $d_1^a = 6.17 \times 10^{-1}$ for the first solution and $d_1^b = 8.50 \times 10^{-5}$ for the second. This illustrates a limitation of direct nonlinear model fitting that is encountered even in very simple noise-free cases. The functions f_{1a} and f_{1b} corresponding to the respective solutions $\hat{x}_a$ and $\hat{x}_b$ are plotted with f_1 as reference in figure 2.

These examples demonstrate that in order to obtain a reliable nonlinear fit to an e-sum, it is necessary to perform both multistart optimization and model selection for the number of terms in the e-sum.

3. Methods based on the Laplace transform and its inverse

The Laplace transform (LT) of a function f is given by

$$\mathcal{L}f = \int_0^\infty f(t)\, \mathrm{e}^{-st}\, \mathrm{d}t = g(s) \tag{6}$$

where s is the complex LT variable.

Determination of the parameters of a linear combination of real decaying exponentials from a measured spectrum $g(s)$ is a special case of the inversion of the LT (Lee *et al* 1993, Schnedermann 1994). In this section, we discuss the ill-posedness of the inverse Laplace transform (ILT) and regularization methods that have been proposed to improve the condition of this inverse problem. We note from the outset that ESA algorithms are applied to time series data rather than exponential spectra. As a consequence, the relevant time series $i(t_l)$ and $\phi(t_l)$ must be transformed into the Laplace domain in order to apply ILT methods, which then numerically evaluate the inverse transform. Owing to the ill-posedness of the ILT, it is difficult to justify the application of the LT to time domain data so that ILT methods may then be applied.

We first illustrate the methods based on the LT. We begin by transforming (3) into the Laplace domain and show how this transformation, followed by discretization, leads to systems of equations that are linear in terms of the unknown parameters. For simplicity, we ignore the normalization (2), and condense the linear coefficient of each exponential into the parameters κ_1^m.

Applying the LT to (3) for $t > 0$ yields

$$\mathcal{L}\phi(t) = \Phi(s) = I(s) \sum_{\tilde{m}=1}^{\tilde{M}} \frac{\kappa_1^{\tilde{m}}}{s + k_2^{\tilde{m}}} \tag{7}$$

where $\Phi(s)$ is the LT of $\phi(t)$ and $I(s)$ is the LT of the input function $i(t)$. A quadrature formula having weights w_l is applied to both $\phi(t)$

$$\Phi(s) = \sum_{l=1}^{L} \mathrm{e}^{-t_l(s-1)} \phi(t_l) w_l \tag{8}$$

and $i(t)$ in order to convert these time series to the LT domain (Bellman 1966, pp 89–90).

The least-squares problem

$$\min_{\kappa_1^{\tilde{m}}, k_2^{\tilde{m}}} \sum_{s=1}^{L} \left[\sum_{\tilde{m}=1}^{\tilde{M}} \frac{s + k_2^{\tilde{m}}}{\kappa_1^{\tilde{m}}} \Phi(s) - I(s) \right]^2 \tag{9}$$

is linear in $k_2^{\tilde{m}}$. The parameters may be estimated by solving the system of normal equations associated with (9) (Bellman 1966, pp 89–90).

This approach is not practical for ESA because

- there is no convenient way of restricting the position of the poles $s = -k_2^{\tilde{m}}$ to the real axis. With no such restrictions in place, this method will tend to fit oscillatory modes to noise in the data.
- no allowance is made in the formulation for noise modelling, model mismatch or representation of noise in the Laplace domain.

Several related approaches suffer from similar drawbacks. These include methods due to Prony (1795), Steiglitz (1977) and Osborne and Smyth (1991).

Many other methods have been proposed for inverting the LT. Several inversion formulae for the LT have been proposed, but they are not considered useful for most applications, since these formulae are developed without considering the ill-posedness of the problem and sensitivity to noise (Widder 1941, Bellman 1966, Ramm 1986, Dong 1993). Many formulae involve differentiation, or are impractically complex (Dong 1993). The approximate inversion formula of Miller and Guy (1966), which is based on Jacobi polynomials, is more practical, but ignores the ill-posedness of the problem. Analyses that have considered ill-posedness, such as those due to McWhirter and Pike (1976) and Varah (1983), are restricted in their application by the respective requirements that the disturbance in the signal is representable in the image space of the Laplace transform, and that the noise level of the analyzed time series is low (Dong 1993).

In order to effect the inverse Laplace transform (ILT), practical methods always place restrictions on the class of s-domain functions that can be inverted (Ang *et al* 1989). Owing to the ill-posedness of the ILT, the most restrictive inversion method available should generally be used to invert a particular function.

The most commonly employed techniques for numerical inversion of the LT involve regularization of the inverse problem. Consider a discretization of the integral equation (6):

$$\mathbf{K}\,\mathbf{f} = \mathbf{g} \tag{10}$$

where **K** is an $L \times L$ matrix and L is the dimension of vector g (Varah 1983). When **K** is ill-conditioned, a regularized solution may be obtained using several methods, an extensive survey of which appears in Varah (1983). In general, ILT methods are not couched in the framework of parameter estimation, but rather in one of estimating a smooth time-series **f** to satisfy the linear system (10) subject to a smoothness constraint. As a consequence, such methods are not appropriate for ESA.

4. Exponential spectral analysis

Gardner *et al* (1959) and Provencher (Provencher 1976, Provencher and Dovi 1979) developed methods of analysing exponential decay curves that recast the ESA problem as one of solving a Fredholm integral equation. The method of Provencher, which includes a hypothesis test for estimating the number of exponential components $\tilde{M}$, is of practical utility in the presence of noise, and has been applied to several practical problems relating to chemical relaxation, relaxographic magnetic resonance imaging (Lee *et al* 1993, 1999), activation energy spectra and ligand binding kinetics (Provencher and Dovi 1979). Variants of the method allow smoothness criteria to be applied to the spectrum by means of regularization (Provencher and Dovi 1979). This formulation is very attractive, since many of the methods available for the inversion of noisy and ill-conditioned integral equations may be used to solve the problem (Provencher 1982). The algorithm presented in [4], for example, uses truncated SVD regularization (Varah 1983). As a consequence, the number of components in the exponential sum may be conveniently estimated from the singular values of the system of linear equations associated with the integral equation.

The methods of Provencher have several drawbacks in the present context:

(i) These methods are derived for decomposing exponential sums, rather than the convolution of exponential sums with an input function. Without modification, a preliminary deconvolution step would be required to solve for the parameters of (3). This process is itself typically ill-posed.

(ii) The input data are smoothed before processing using an autocorrelation formula. The influence of this smoothing process on the solutions obtained is not obvious and has not been analysed.

(iii) When smoothness constraints are placed on the spectrum, the determination of the level of smoothness is somewhat arbitrary in the absence of specific prior knowledge relating to the form of the spectrum. Also, simultaneous imposition of smoothness and non-negativity constraints on the exponential rate constants can lead to difficulties.

Cunningham and Jones presented an ESA method in which a model of the form (3) was fit to time-activity data (Cunningham and Jones 1993). Instead of optimizing over the k_1 and the k_2, the k_2 were preselected and a non-negative least-squares (NNLS) algorithm was used to solve for the linear parameters k_1. A total of $\tilde{M} = 100$ exponential terms, or modes, were chosen. The rate constants of this set were obtained by regular sampling along a log axis over a predefined range. Apart from the advantage of not needing to know *a priori* how many terms are present in an e-sum, spectral estimation can avoid the problem of sensitivity with respect to the starting estimate of the parameter vector. In addition, since the spectral coefficients are linear parameters, parameter sets may be interpreted as histograms describing the contributions of each mode. Since the selection of rate constants of the basis set remains constant for a specific application, the comparison of the results of repeated experiments is thus facilitated.

The most significant problem with the Cunningham and Jones ESA (CJESA) implementation is the extreme sensitivity of the spectral parameter estimates to noise. This is due to the parameter redundancy present in e-sums, which occurs even when the number of basis function $\tilde{M}$ is as small as two, as shown in section 2. When $\tilde{M} = 100$, the variability that the parameter redundancy introduces makes it extremely difficult to meaningfully compare spectral coefficients even at high signal-to-noise ratios (SNRs).

5. Parsimonious exponential spectral analysis

The objective of this work is to refine the method of Cunningham and Jones by means of a rational selection of the exponential modes that make up the spectral basis. We describe the selection as rational and parsimonious rather than optimal, because parameter redundancy cannot ever be eliminated entirely. This is why we choose to select the model of lowest complexity that is able to adequately describe the data.

Owing to the linear dependence between the basis elements, the solution obtained will always constitute a compromise between goodness of model fit and parameter variance. The way in which this compromise is attained is application-dependent and must be determined by the user of the ESA algorithm.

The principal strategy of PESA is to find a small set of exponentials that are able to approximate a large set of exponential basis functions, such as that employed in CJESA. While reduced basis dimension increases approximation error, it decreases parameter variance. Compromise between these two cost measures may be achieved by adjusting the number of basis functions so that the specifications of a particular application may be met.

We now elaborate on the characteristics that a spectral representation should exhibit:

(i) Low approximation error, which reflects an accurate fit.
(ii) Low bias, which suggests that parameter redundancy has been reduced to an acceptable degree.
(iii) Tight Cramér–Rao lower bound (CRLB) at the solution. This is important when the spectral coefficient vectors from several experiments are to be combined into a mean vector for comparison with means obtained from other sets of experimental results.

In the PESA algorithm described in the next section, a compromise between these criteria is achieved via a model selection process.

In order to perform ESA, the following are required:

(i) A set of time series to which the model is to be fitted.
(ii) A set of forcing (input) function time series.
(iii) A range of k_2 over which the exponential basis must be generated.
(iv) The number of basis elements $\tilde{M}$.

While the choice of a range of rate constants is easily justified in terms of *a priori* knowledge regarding feasible values for a specific application, an arbitrary choice of $\tilde{M}$ introduces difficulties. Specifically, when $\tilde{M}$ is too high, the parameter estimates of the spectral representation may not be unique or may lack robustness in the presence of noise. When $\tilde{M}$ is too low, the time series cannot be modelled with sufficient accuracy.

Application of the singular value decomposition (SVD) to the spectral basis employed by Cunningham and Jones yields relatively few (four to six) significant singular vectors (Maltz 2000). This illustrates the large degree of redundancy in the basis. It also provides an orthogonal basis in terms of which e-sums may be approximated. Unfortunately, the coefficients of this basis do not correspond to physically meaningful quantities. In addition,

the basis elements generally exhibit negative excursions. This is inconvenient when modelling non-negative physical quantities such as concentrations and photon count values, since additional non-negativity constraints must be applied in the data space as well as in the parameter space.

Affine transformations have been applied to these orthogonal functions to allow the response of single-compartment models to be represented using non-negative basis functions, having non-negative coefficients (Sitek 1999, Maltz 2001). In such a formulation simple interval bound constraints on the parameters are sufficient to ensure the non-negativity of the model output. The drawback of these methods is that the coefficients do not generally quantify underlying specific physical or biological processes.

5.1. Parsimonious basis selection

We wish to find a small basis of exponential functions that can approximate the elements of a redundant basis set over a specified range of rate constants $k_2^m \in \left[k_2^1, k_2^{\tilde{M}}\right]$. We begin by forming the basis $\mathbf{E}_{\tilde{M}}$ proposed by Cunningham and Jones, based on the general ESA matrix

$$\begin{aligned}\mathbf{E}_M &= [\mathrm{e}_1 \quad \mathrm{e}_2 \quad \cdots \quad \mathrm{e}_M] \\ &= \begin{bmatrix} k_2^1 \mathrm{e}^{-k_2^1 t_1} & k_2^2 \mathrm{e}^{-k_2^2 t_1} & \cdots & k_2^M \mathrm{e}^{-k_2^M t_1} \\ k_2^1 \mathrm{e}^{-k_2^1 t_2} & k_2^2 \mathrm{e}^{-k_2^2 t_2} & \cdots & k_2^M \mathrm{e}^{-k_2^M t_2} \\ \cdots & \cdots & \cdots & \cdots \\ k_2^1 \mathrm{e}^{-k_2^1 t_L} & k_2^2 \mathrm{e}^{-k_2^2 t_L} & \cdots & k_2^M \mathrm{e}^{-k_2^M t_L} \end{bmatrix}\end{aligned} \tag{11}$$

with $\tilde{M} = 100$ and where the k_2 are regularly spaced on a log axis. While we work with the basis used by Cunningham and Jones for concreteness, we do this without loss of generality, since we require only that $\mathbf{E}_{\tilde{M}}$ be column rank deficient.

As we wish to approximate $\mathbf{E}_{\tilde{M}}$ with a matrix $\mathbf{E}_M$ with $M \ll \tilde{M}$ by a process of model selection, it is helpful to obtain a starting estimate for M.

This estimate will have some dependence on the forcing function with which the exponential basis is convolved in (3). Let us assume that each column of $\mathbf{E}_{\tilde{M}}$ is convolved with a time series representing a typical forcing function for a specific application, to form the respective columns of a matrix of identical dimension $\mathbf{C}_{\tilde{M}}$.

Application of the SVD to $\mathbf{C}_{\tilde{M}}$ yields *inter alia* the L left singular vectors $\mathbf{u}_m$. Each of the N time series to be analyzed, $\phi^n(t_l)$, may be approximated in terms of this orthogonal basis using the first M singular vectors (assuming $L \geqslant M$) in the series as

$$\hat{p}_M^n(t_l) = \sum_{m=1}^{M} \mu_{mn} u_{lm} \tag{12}$$

where u_{lm} is the lth element of $\mathbf{u}_m$.

The cumulative energy in the approximation up to the Mth term, normalized by the total energy in $\phi^n(t_l)$, is

$$Z_M^n \triangleq \left[\sum_{m=1}^{M}\sum_{l=1}^{L^n} \hat{p}_M^n(t_l)^2\right] \Bigg/ \sum_{l=1}^{L^n} \phi^n(t_l)^2. \tag{13}$$

Here L^n is the length of the nth time series $\phi^n(t_l)$.

As criterion for choosing that value of M at which the approximation error and noise contain similar energy, we propose the metric M_0, defined as

$$M_0^n \triangleq \max\left\{M \;\middle|\; \frac{1}{1 - Z_M^n} \leqslant \mathrm{SNR}\right\} \tag{14}$$

where SNR is the signal-to-noise ratio of the data. This estimate is derived assuming that singular vectors are used as approximating bases. We may, in general, expect the number of exponential basis functions needed to approximate the data to be greater than or equal to M_0, since the exponential basis is neither orthogonal, nor an optimal approximating basis for the representation of the data. To err on the side of caution, we employ the more conservative starting value for M in the model selection process

$$M_-^n = \max\left\{ M \;\middle|\; \frac{1}{1 - Z_M^n} \leqslant \mathrm{SNR}/10 \right\}. \tag{15}$$

During each step of model selection, the value of M is incremented, a candidate parsimonious basis set is calculated and a spectral model of dimension M is fitted to the dataset. The robustness of the solution in the presence of noise is then tested by refitting the model to the time series generated from the solution and containing appropriately modelled noise. This process of estimation over many noise-realizations is employed to determine whether solutions obtained for a particular time series, using this estimator, are biased (perhaps because of redundancy in the basis). Usually, the term 'bias' refers to the difference between the expected value of the parameter estimate and its true value. Here, 'bias' refers to the difference between the expected value of the parameters (estimated using simulation of the model) and the estimate of the solution obtained using this model of dimension M.

In parameter estimation problems where the existence of a single unique solution is guaranteed, theoretical bounds on parameter variance such as the Cramér–Rao lower bound are sufficient to evaluate the robustness of this solution. In the case of fitting parameters to the ESA basis, simulation is necessary in order to ensure that the solution does not 'jump' between different basins-of-attraction within the hypersurface of the cost function.

Figure 3 serves as illustration of how simulation may be used to determine whether multiple solutions are present. Here, 100 nonlinear fits of the function f_1 introduced in figure 1 are performed. Gaussian white noise of variance 0.25 is added to f_1 before parameter estimation is performed. In figure 3(a), the optimization algorithm is initialized with $x_0 = \left[k_{10}^1\, k_{10}^2\, k_{20}^1\, k_{20}^2\right] = [1\,2\,3\,4]$. The mean of the parameter estimates coincides closely with the true solution, and the variance of the estimates tends asymptotically towards the CRLB as is theoretically expected of an efficient estimator. In figure 3(b) the parameter estimator is randomly initialized at either $x_0 = [0\,0\,0\,0]$ or $x_0 = [1\,2\,3\,4]$. The presence of a suboptimal local maximum of the likelihood function introduces large bias in the estimates, and the parameter variance significantly exceeds the CRLB as predicted, assuming a single (true) solution.

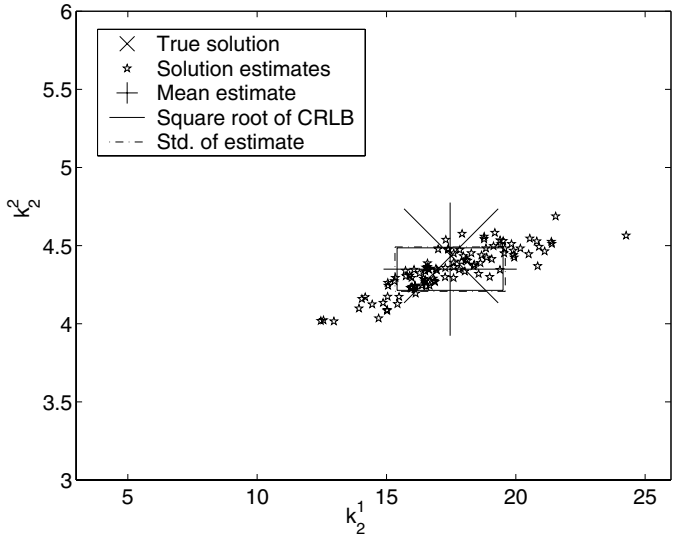
If the solutions from the simulation exhibit acceptably low bias, and if the approximation accuracy is acceptable, model selection terminates. Often it is desirable to precalculate the $\mathbf{C}_M$ over a certain range before the data analysis ensues. For this purpose, we propose that basis generation be terminated at a value of M greater than

$$M_+^n = \max\left\{ M \;\middle|\; \frac{1}{1 - Z_M^n} \leqslant 10 \times \mathrm{SNR} \right\}. \tag{16}$$

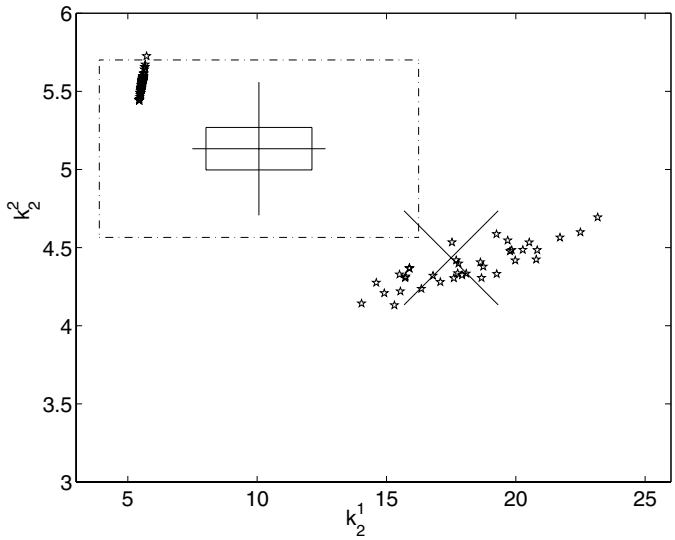
In practice, since the values of M_-^n and M_+^n depend on the particular time series $\phi^n(t_l)$, it is necessary to invoke PESA with the algorithm parameters

$$M_- = \min_{n=1,2,\ldots,N} M_-^n \tag{17}$$

$$M_+ = \max_{n=1,2,\ldots,N} M_+^n. \tag{18}$$



(a)



(b)

Figure 3. Illustration of the way in which the presence of multiple solutions may be revealed via simulation. (a) The parameter estimation algorithm efficiently estimates the true solution. However, in (b), owing to differences in algorithm initialization parameters, some estimates fall within a suboptimal local maximum of the likelihood function. By examining bias and comparing parameter variance to the CRLB that applies at the true solution, the presence of the spurious solutions is revealed.

For each value of M, we find the rate constants that define the candidate basis set by solving

$$\min_{\mathbf{k_2},\boldsymbol{\mu}} \sum_{l=1}^{L} \sum_{\tilde{m}=1}^{\tilde{M}} \left[\tilde{k}_2^{\tilde{m}} \, \mathrm{e}^{-\tilde{k}_2^{\tilde{m}} t_l} - \sum_{m=1}^{M} \mu_{\tilde{m}}^{m} \, \mathrm{e}^{-k_2^m t_l} \right]^2 \qquad t_l \in [0, T] \tag{19}$$

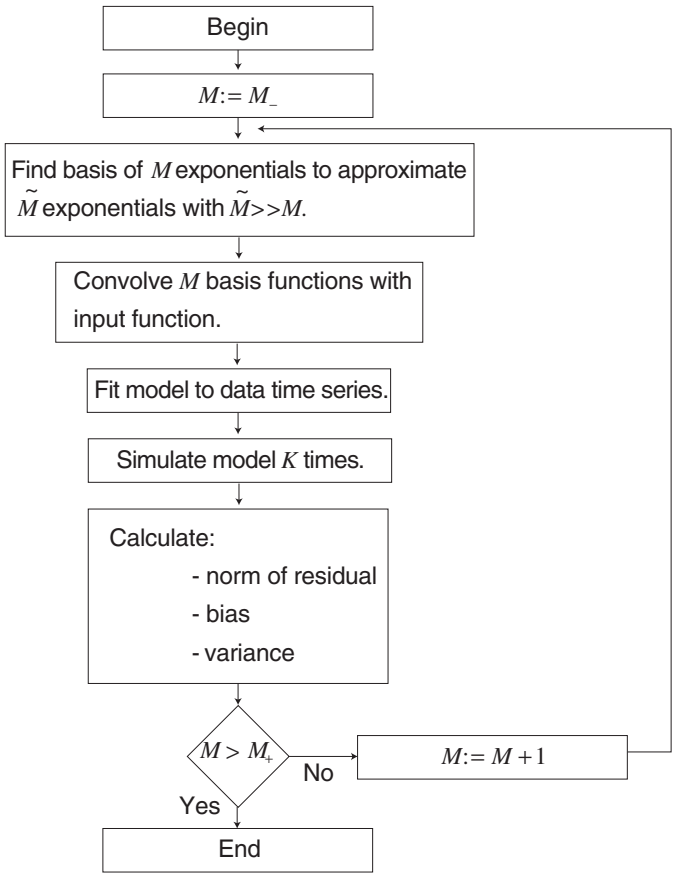


Figure 4. Flowchart of the model selection process in the PESA algorithm for a single time series and forcing function pair. The symbol M denotes the current basis dimension, while the limits M_-, M_+ are defined in (15) and (16), respectively. A more detailed flowchart of the algorithm (for multiple time series) appears in figure 5.

under the constraints $\mu_{\tilde{m}}^m \geqslant 0, k_2^m \geqslant 0, m = 1, 2, \ldots, M$. Here $\boldsymbol{\mu}$ and $\mathbf{k_2}$ are vectors containing the $\mu_{\tilde{m}}^m$ and k_2^m, respectively. In the solution of this problem, only $\mathbf{k_2}$ is of interest.

The minimization problem (19) may be solved using any nonlinear NNLS algorithm. We employ a bound-constrained quasi-Newton method that makes use of analytically derived first derivatives provided by the user (NAG Fortran routine E04KYF, Numerical Algorithms Group, Oxford, UK).

5.2. Model selection algorithm

The primary motivation for the use of a parsimonious basis is to reduce the variability of the coefficients of the spectral model. This is achieved, ideally, without introducing model approximation error that is of greater magnitude than the signal noise. In this section, we describe a procedure for choosing the dimension of the basis to adequately satisfy these criteria. A flowchart of this procedure as applied to a single time series and its corresponding forcing function appears in figure 4.

In the discussion that follows, we assume that the data consist of N time series $\phi^n(t_l)$; $n = 1, \ldots, N$; $l = 1, 2, \ldots, L^n$, each of which may be modelled as the convolution of an input function $i^n(t_l)$ with an e-sum, as described in (3).

We also assume that a set of $\mathbf{E}_M$, usually with M in the range M_- to M_+, has been generated through the solution of (19). The columns of each $\mathbf{E}_M$ have been convolved with the input functions $i^n(t_l)$ to form the N matrices $\mathbf{C}_M^n$.

For a specific value of M and n, the spectral coefficients are obtained via solution of the NNLS problem

$$\hat{\boldsymbol{\mu}}_M^n = \arg\min_{\boldsymbol{\mu}} \left\| \mathbf{D}_n \left(\boldsymbol{\phi}^n - \mathbf{C}_M^n \boldsymbol{\mu} \right) \right\| \tag{20}$$

$$\mu_m \geqslant 0 \qquad m = 1, 2, \ldots, M \tag{21}$$

where the μ_m are the elements of the spectral coefficient vector $\boldsymbol{\mu}$, and $\boldsymbol{\phi}^n$ contains the nth data time sequence. $\mathbf{D}_n$ is a weighting matrix

$$\mathbf{D}_n \triangleq \mathbf{W}^{-1/2} = \operatorname{diag}(1/\sqrt{w_1}, 1/\sqrt{w_2}, \ldots, 1/\sqrt{w_{L^n}}) \tag{22}$$

where L^n is the length of the nth time series. Inclusion of this matrix allows a weighted least squares solution to be obtained for a Gaussian noise model having a diagonal covariance matrix $\mathbf{W} = \Lambda$.

Where the time sampling of the data is irregular, or where the basis functions are not sampled at the same points as the data, we perform interpolation onto a regularly sampled axis in order to effect the convolution. Each time series and its corresponding input function is interpolated onto a time axis from $t = t_1^n$ to $t = t_L^n$, the final sample time, using a uniform sample spacing

$$\Delta t^n = \min_{l=1,2,\ldots,L^n-1} \left(t_{l+1}^n - t_l^n \right) \tag{23}$$

where t_l^n are the L^n original sample times for the nth time series. The number of samples on the nth interpolating time axis is denoted $\tilde{L}^n$.

Interpolation followed by convolution yields an $\tilde{L}^n \times M$ matrix, the columns of which are reinterpolated onto the original time axis to yield the $L^n \times M$ matrix $\mathbf{C}_M^n$. The latter is required to solve (20).

Using $\hat{\boldsymbol{\mu}}_M^n$, we may reconstruct the time series

$$\hat{\boldsymbol{\phi}}_M^n = \mathbf{C}_M^n \hat{\boldsymbol{\mu}}_M^n. \tag{24}$$

As model selection ensues, it is important to verify that each set of spectral coefficients $\hat{\boldsymbol{\mu}}_M^n$ constitutes a robust spectral estimate for the time series that it models. As alluded to earlier, we propose to obtain some measure of the robustness of each solution via repeated estimation of the spectral coefficients for $\hat{\boldsymbol{\phi}}_M^n$ under simulated noise conditions. The effectiveness of this technique is strongly dependent on the accuracy of the noise model used to create the simulated time series. It is especially important that the SNRs of the simulated and real data are similar.

We generate the K simulated time series

$$\boldsymbol{\phi}'^n_k = \hat{\boldsymbol{\phi}}_M^n + \boldsymbol{\eta}_k \tag{25}$$

where $\boldsymbol{\eta}_k$ is the kth realization of a noise process. Application of PESA to these time series yields K solution vectors $\tilde{\boldsymbol{\mu}}_M^{n,k}$. The bias, assuming that $\hat{\boldsymbol{\mu}}_M^n$ is the true solution, is estimated as

$$\mathbf{b}_M^n = \frac{1}{K} \sum_{k=1}^{K} \tilde{\boldsymbol{\mu}}_M^{n,k} - \hat{\boldsymbol{\mu}}_M^n. \tag{26}$$

When forming a metric of overall bias, it is important to discount those large percentage biases that are observed when the underlying parameter does not contribute significantly to the modelled time series. To address this problem, we form the time series

$$\beta_M^n = \mathbf{C}_M^n \left| \mathbf{b}_M^n \right| \tag{27}$$

where $|\cdot|$ denotes the element-by-element absolute value operator. We then evaluate the energy in the absolute bias expansion as a fraction of the total energy in the time series

$$Y_M^n = \sum_{l=1}^{L^n} \beta_M^n(t_l)^2 \Bigg/ \sum_{l=1}^{L^n} \phi^n(t_l)^2 \tag{28}$$

where $\beta_M^n(t_l)$ is the lth element of β_M^n. Our measure of the overall bias for a particular value of M over all n is defined as

$$B_M \triangleq \frac{1}{N} \sum_{n=1}^{N} Y_M^n. \tag{29}$$

If, at a particular value of M, B_M is suitably small, the approximation error must be evaluated to establish the accuracy of the model. For a weighted fit, the residual norm in (20) evaluated at the attained minimum is an appropriate metric of approximation error. The quality of the fit is easier to gauge, however, using the criterion

$$R_M = \frac{1}{N} \sum_{n=1}^{N} \frac{\left\| \mathbf{D}_n \left(\phi^n - \mathbf{C}_M^n \hat{\boldsymbol{\mu}}_M^n \right) \right\|^2}{\left\| \mathbf{D}_n \phi^n \right\|^2}. \tag{30}$$

Figure 5 contains a detailed flowchart of the model selection procedure and illustrates the way in which N output time series and forcing function pairs are accommodated. The algorithm consists of three nested main loops. The outer loop is the model selection loop over the basis dimension M. The loop over all N time series is nested below the M-loop. The K simulations form the innermost loop. The M-loop terminates when a suitable compromise between bias and approximation error is found.

Often, a single model is desired to represent the 'mean' result of N experiments. This is especially important when the results of two or more sets of experiments are to be compared. The meaningfulness of a mean parameter vector

$$\bar{\boldsymbol{\mu}}_M = \frac{1}{N} \sum_{n=1}^{n} \hat{\boldsymbol{\mu}}_M^n \tag{31}$$

is dependent on the variability of its elements. We address this issue in the following section.

5.3. Variance of the solution

Let us denote the value of M chosen at the end of the above algorithm by I, and the solution $\hat{\boldsymbol{\mu}}_I^n$ for a specific time series ϕ^n by $\hat{\mathbf{a}}$.

The Cramér–Rao lower bound is a measure of the mean curvature of a likelihood function at a particular solution point and is a bound on the variance of the solution. Where this bound is loose, the solution may be unacceptably sensitive to noise. Calculation of the CRLB provides a convenient means of assessing the robustness of spectral estimates.

An estimator that achieves the CRLB bound is termed an efficient estimator. The CRLB for unbiased estimates is given by

$$\mathrm{Var}\left[\hat{a}_i(\mathbf{R}) - A_i\right] \geqslant (\mathbf{J}^{-1})_{ii} \qquad i = 1, 2, \ldots, I \tag{32}$$

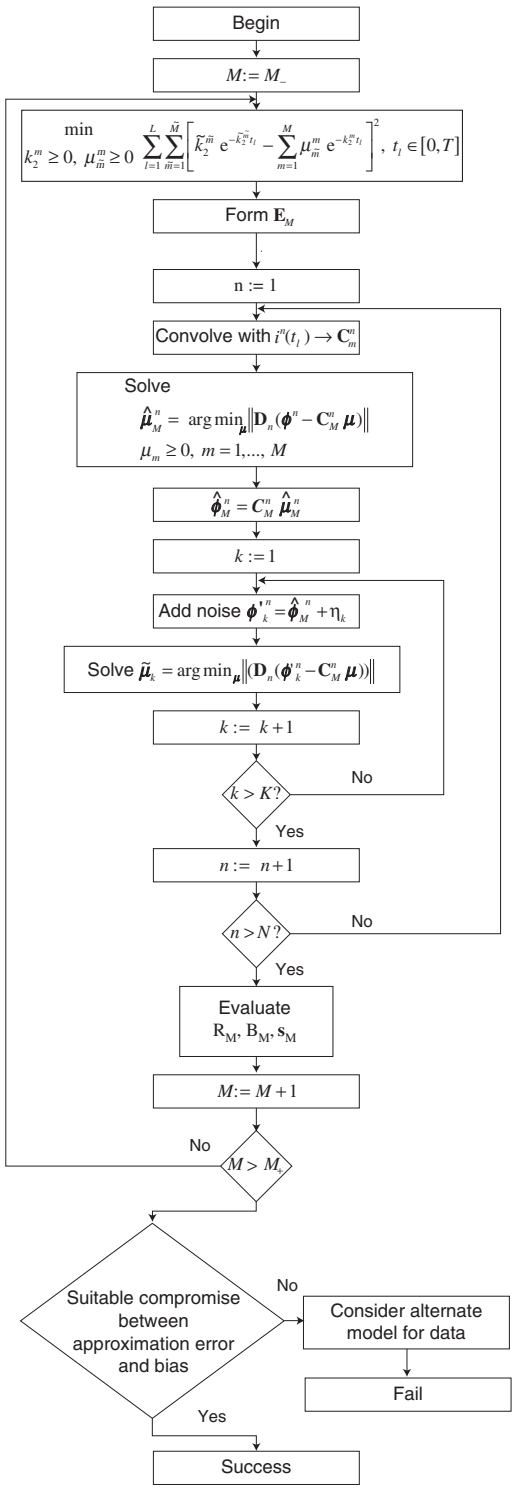


Figure 5. Flowchart of the PESA algorithm.

where $\hat{a}_i$ is the ith element of $\hat{\mathbf{a}}$, A_i is the corresponding element of the true solution and $(\mathbf{J})_{ii}$ is the ith diagonal element of $\mathbf{J}^{-1}$. The latter is the inverse of the Fisher information matrix (FIM) whose elements are defined as

$$J_{ij} \triangleq -E\left[\frac{\partial^2 \log(p_{\mathbf{r}|\mathbf{a}}(\mathbf{R} \mid \mathbf{A}))}{\partial A_i \partial A_j}\right]. \tag{33}$$

Here the probability of the data vector $\mathbf{R}$ given the nonrandom parameter set $\mathbf{A}$ (containing the I elements A_i) is given by $p_{\mathbf{r}|\mathbf{a}}(\mathbf{R} \mid \mathbf{A})$. This notation is chosen to be consistent with that used by Van Trees (1968).

Since the FIM (33) may be considered as an approximation to the Hessian of the log-likelihood function $\log p_{\mathbf{r}|\mathbf{a}}(\mathbf{R} \mid \mathbf{A})$, the CRLB may be interpreted as the mean curvature of this function at the solution. The greater the curvature at a solution, the greater confidence we have that the solution is a unique maximizer of the likelihood within its 'basin of attraction', and the greater is its robustness against noise.

Where multiple solutions are present, an efficient estimator will not achieve the CRLB, owing to the fact that the estimator will converge to an alternate solution some of the time.

The CRLB as defined in (32) and (33) is not directly applicable to constrained problems since the expectation of the curvature is taken over directions that are infeasible owing to the imposed constraints. Gorman and Hero have derived a constrained CRLB (cCRLB) for cases where the parameters are subject to multiple smooth nonlinear inequality constraints (Gorman and Hero 1990). They demonstrate that orthant or general support constraints lead to a constrained FIM that is a dimension-reduced version of the original FIM. In the spectral estimation problem, the presence of non-negativity constraints on the parameters will produce maximum likelihood estimates in which some spectral coefficients will be equal to zero. When the gradient of the likelihood function is positive for parameters at their bound, we believe it is justifiable to consider these variables to be constrained to equal their bound. As shown in Gorman and Hero (1990), the rows and columns of the unconstrained FIM may be deleted to form the constrained FIM.

We base the cCRLB on the constrained FIM given by

$$\begin{aligned} J_{ij} &= -E\left[\frac{\partial^2 \log(p_{\mathbf{r}|\mathbf{a}}(\mathbf{R} \mid \mathbf{A})]}{\partial A_i \partial A_j}\right] \\ i, j &= \left\{k \mid A_k > 0, k \mid A_k = 0, \left.\frac{\partial \log(p_{\mathbf{r}|\mathbf{a}}(\mathbf{R} \mid \mathbf{A}))}{\partial A_k}\right|_{A_k=0} \geqslant 0\right\}. \end{aligned} \tag{34}$$

Thus defined, (34) represents the projection of the FIM onto the subspace of unconstrained variables.

Let us consider the mean parameter vector $\bar{\boldsymbol{\mu}}_M$ defined in (31). Assuming that each constituent of the mean represents an unbiased, efficient estimate, the standard deviation of $\bar{\boldsymbol{\mu}}$ achieves the bound

$$\mathbf{s}_M = \sqrt{\frac{1}{N^2}\sum_{n=1}^{N} \mathrm{cCRLB}_M^n} \tag{35}$$

where the vector cCRLB_M^n is the constrained CRLB at the solution $\hat{\boldsymbol{\mu}}_M^n$, and where the square-root is applied element-by-element.

This lower bound provides us with a measure of the variability in the mean spectral vector that is due to the intrinsic properties of the likelihood function at the N individual solutions. These uncertainty estimates facilitate the statistically meaningful comparison of mean vectors obtained from several different sets of experiments.

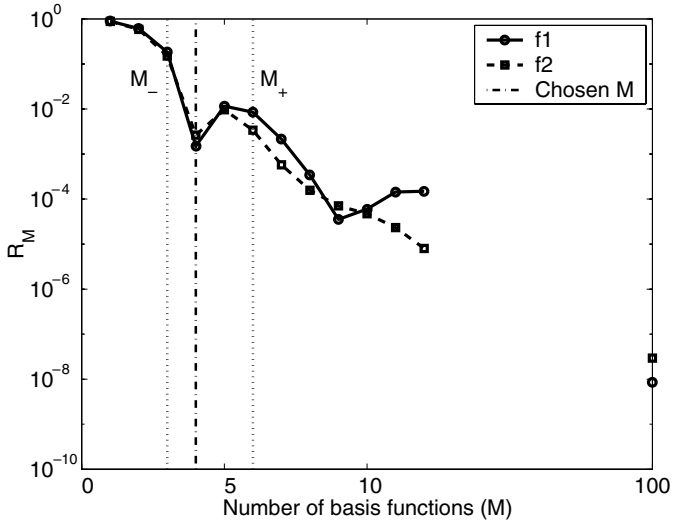


Figure 6. Approximation error metric R_M versus M.

Table 1. Parameters of exponential basis to be approximated using parsimonious basis set for the simulated data application.

Parameter	Value	Unit
T	1	s
L	125	Time samples
$\tilde{M}$	100	Unoptimized bases
$\tilde{k}_2^{\tilde{m}}$	0.001–1000 Log spaced	s^{-1}
K	100	Noise realizations

6. Algorithm evaluation

6.1. Application to synthetic data

We apply PESA to the e-sums f_1 and f_2 described in section 2 using the parameters contained in table 1. To illustrate the generality of the approximation, we choose to approximate an exponential basis that spans a 6-decade range in rate constant. We also normalize the time interval over which the time series are approximated to $t \in [0, 1]$.

Gaussian noise of zero mean and variance of 0.0625 is added to the two functions. This noise level corresponds to an SNR ≈ 400. This high SNR was chosen to illustrate that multiple solutions to the spectral problem are encountered at noise levels far lower than those typically encountered in most physical measurement situations.

Following the procedure outlined above, we calculate $M_- = M_-^1 = M_-^2 = 3$ and $M_+ = M_+^1 = M_+^2 = 6$ as the estimated interval for the optimal value of M. The values at which the approximation errors are estimated to be closest to the noise level are $M_0^1 = 5$ and $M_0^2 = 3$.

For the purpose of illustration, we extend the range of M used in this example and vary M from 1 to 12 basis functions. Figure 6 illustrates the behaviour of the approximation error metric R_M versus M, while figure 7 shows the bias metric B_M versus M. R_M will not, in

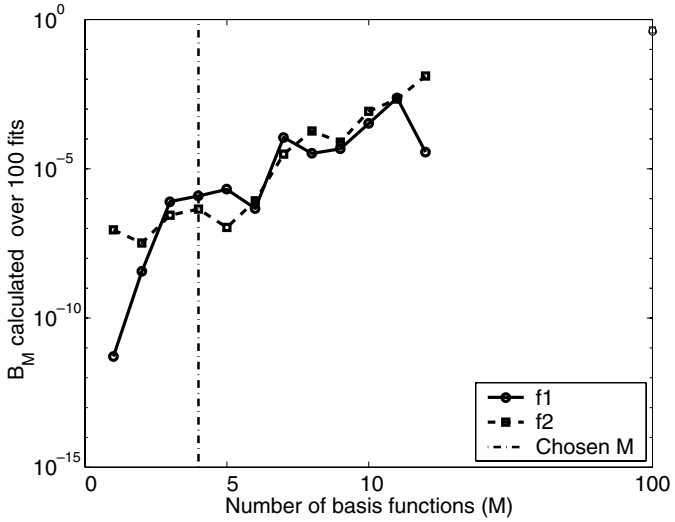


Figure 7. Maximum fractional bias metric B_M versus M. The value of B_M increases with M owing to the tendency of the estimator to model noise present in the data when redundant parameters are present.

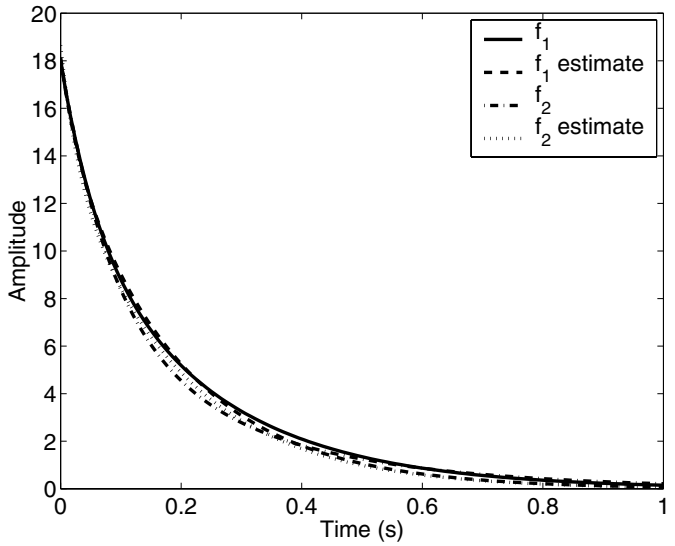


Figure 8. Functions f_1 and f_2 are shown together with their PESA fits for $M = 4$.

general, be a monotonically decreasing function of M because some small dimension basis sets possess elements particularly well suited to approximating particular functions.

Examining these criteria simultaneously, we find that the best compromise between bias and approximation error is achieved at $I = M = 4$. For $M > 6$, the values of B_M increase significantly. The CJESA values are shown at $M = 100$ in both figures. As anticipated, while the approximation error is lower for CJESA, B_M is much greater. The fit for $I = 4$ appears in figure 8.

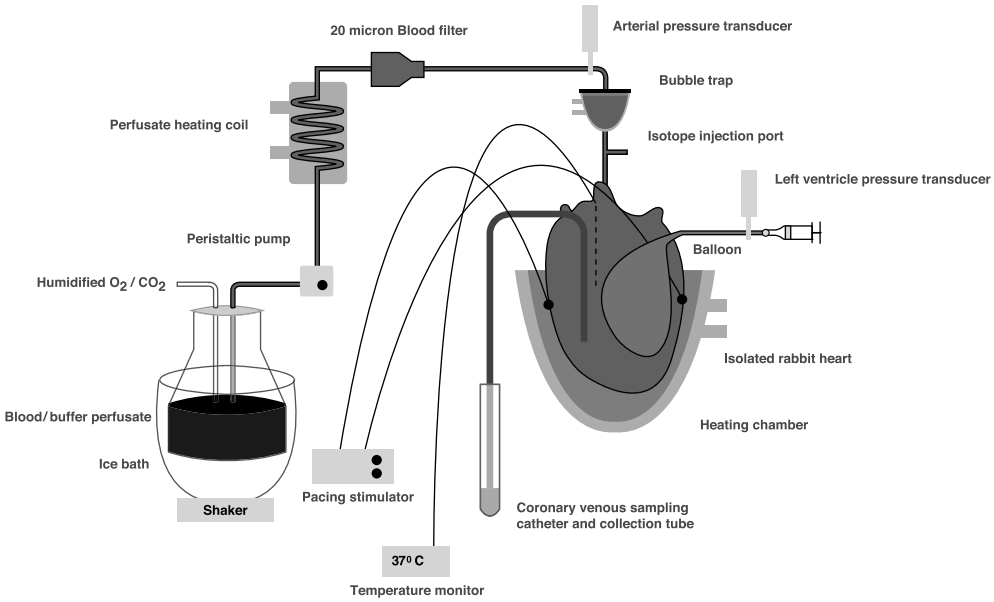


Figure 9. Experiments on the isolated rabbit heart are performed within an apparatus that continuously perfuses the heart. Isotopes are injected just above the aortic cannula. The radiotracer activity in the venous output is measured in a well counter.

Total computation time for PESA applied to both f_1 and f_2 over this range of M is under 3 min on a Pentium III 850 MHz processor. This includes the time taken to form the PESA basis sets for $M = 3, 4, 5$ and 6.

Since $N = 1$ in this evaluation, we do not need to evaluate $\mathbf{s}_M$; however, it is interesting to compare the stringency of the bound on the variance that exists for PESA at $M = 4$ and for CJESA at $\tilde{M} = 100$. To do this we use the metric

$$W_M^n \triangleq \frac{\sqrt{\text{cCRLB}^n}}{\hat{\boldsymbol{\mu}}_M^n} \times 100 \tag{36}$$

where the square root and division operations are performed element-by-element on the vector arguments. We find for f_1 that the maximum element of W_4^1 is 15.4% versus 328% for CJESA. For f_2 these respective values are 10.0% in W_4^2 and 784% in W_{100}^2. The bias and variability of the CJESA coefficients are therefore unacceptably high in comparison with those obtained using PESA.

6.2. Application to radiotracer time-activity modelling

We now illustrate the application of PESA to the modelling of the TACs of radiotracers in the venous outflow of isolated rabbit hearts mounted in a continuous perfusion apparatus. The purpose of this study is to compare the retention and washout characteristics of a new tracer for the determination of myocardial perfusion (^{125}I-iodorotenone) with those of a tracer in widespread clinical use (^{99m}Tc-sestamibi).

The details relating to the acquisition of these data have been reported previously by Marshall *et al* (2001). Briefly, isolated rabbit hearts ($N = 25$) in the perfusion apparatus shown in figure 9 are perfused with a solution containing electrolytes, bovine erythrocytes and serum albumin. During the continuous perfusion process, three radiotracers are simultaneously injected as a compact bolus into a port just above the aortic cannula. The rationale behind

the sudden, brief injection of the tracer is to elicit the impulse response of the system to the three tracers, each of which exhibits a different biokinetic behaviour. One of the tracers, ^{131}I-albumin serves as a reference tracer for the blood flow, since it is not preferentially retained by the myocardium and hence possesses similar biokinetics to blood. The ^{131}I-albumin time-activity in the venous outflow may consequently be used to model the blood input function to the myocardium. The diffusible tracer fractional venous appearance rate $h_d^n(t)$ for each tracer is modelled as consisting of a component that is not retained by the myocardium plus several components whose venous appearance is delayed, owing to their retention inside myocardial compartments

$$h_d^n(t) = h_r^n(t) * i^n(t) \tag{37}$$

$$= h_r^n(t) * \left[c_0^n \delta(t) + \sum_{m=1}^{M} \frac{c_m^n}{\tau_m} \mathrm{e}^{-t/\tau_m} \right]. \tag{38}$$

Here, $h_r^n(t)$ is the reference tracer (^{131}I-albumin) venous appearance function, the $\tau_m = 1/k_2^m$ are the time constants and the c_m are the spectral coefficients. The index n corresponds to the experiment in which all three tracers are perfused into the nth heart.

6.2.1. Basis selection. In order to employ basis dimension selection criteria (15) and (16), a noise model of the TAC data is needed. The measurement uncertainty of each TAC sample is provided with the experimental data, having been estimated from the Poisson distribution of the photon count data, as well as from the errors introduced in the process of separating the counts contributed by each of the tracers from the aggregate photon counts recorded by the well counter. This is achieved through least-squares fitting and subsequent discrimination of the energy spectra of the tracers (Marshall *et al* 2001). Owing to the presence of multiple sources of uncertainty, and since there is no reason to believe that the noise is correlated over time, a Gaussian model with diagonal covariance matrix is employed.

The SNR is estimated for the ^{125}I-iodorotenone and ^{99m}Tc-sestamibi sets separately as

$$\min_{n=1,\ldots,N} \frac{1}{L^n} \sum_{l=1}^{L^n} \frac{h_d^n(t_l)^2}{\sigma_d^n(t_l)^2}. \tag{39}$$

The process of model selection and spectral analysis requires that the exponential basis be convolved with the $h_r^n(t_l)$, the relevant kernel functions. This is complicated by the fact that each of the 25 input functions is sampled with different, non-uniform time sampling. To address this problem, interpolation onto the axes described in (23) is performed. The convolved series are then reinterpolated onto the original time axes.

The unoptimized exponential basis is chosen to have 100 log-spaced components with time constants in the range [1/60, 190] minutes. These are convolved with the input functions from the set $h_r^n(t_l)$ before the criterion values $M_- = 1$ and $M_+ = 70$ are calculated using both sets of data. The large difference between M_- and M_+ is due to the very high but variable SNR of the data (SNR range: 26 to 2.9×10^4). For this reason, the basis sets are calculated during the model selection process so as to avoid unnecessary precomputation of basis sets of large dimension.

The basis selection parameters are summarized in table 2.

PESA is performed on each of the 25 curves for each experiment and for each of the two retained tracers, yielding 50 non-negative parameter estimate vectors $\mathbf{c}_M^n$ for each value of M. These satisfy the equations

$$\mathbf{c}_M^n = \mathbf{C}_M^n \mathbf{h}_D^n \qquad \begin{array}{ll} M_+ \geqslant M \geqslant M_- & M \in \mathcal{Z} \\ n = 1, 2, \ldots, N. & \end{array} \tag{40}$$

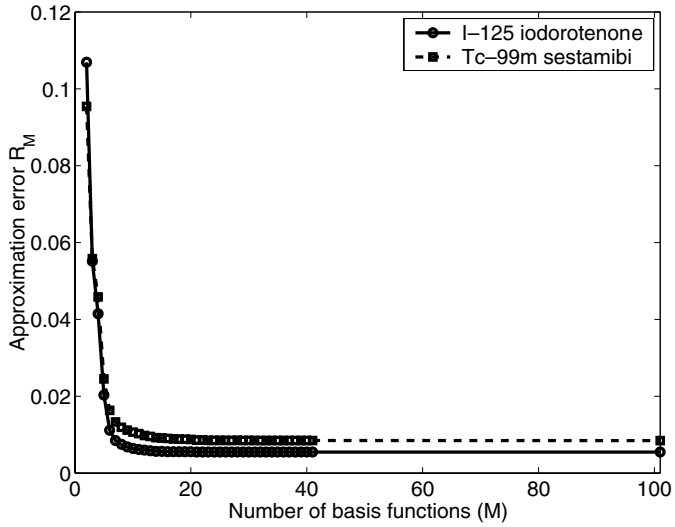


Figure 10. Approximation error metric R_M versus number of basis functions for isolated rabbit heart TACs modelled using PESA. The value for the CJESA estimate appears at $M = 101$.

Table 2. Parameters of exponential basis to be approximated using parsimonious basis set for the isolated rabbit heart application.

Parameter	Value	Unit
T	Variable	s
L	Variable	Time samples
$\tilde{M}$	100	Unoptimized bases
$\tilde{k}_2^m$	1/60–190	min^{-1}
	Log spaced	

Figure 10 shows the approximation error metric R_M versus the number of bases M. It is apparent that large decreases in the approximation error are achieved only up to $M \approx 11$ for both tracer datasets. We examine the behaviour of B_M in figure 11 and find that maximum bias increases significantly for $M > 12$ for ^{125}I-iodorotenone. Consequently, a choice of $I = M = 11$ seems an appropriate parsimonious choice for this dataset. In this plot the sample at $M = 101$ represents the CJESA estimate, which exhibits large bias.

The means of the spectral estimates (over the $N = 25$ datasets) for each of the tracers, along with the standard errors of the means (SEMs), appear in figure 12. The coefficients have been normalized to sum to unity, to aid in interpretation of the coefficients as representing histograms of the modal distribution of time-activity. Immediate inspection reveals that the value of coefficient c_1, corresponding to a small time constant, is higher for ^{99m}Tc-sestamibi than for ^{125}I-iodorotenone. Consequently, a greater fraction of this tracer appears in the venous outflow in the first few seconds after injection. It is also clear from normalized coefficient c_{11} that a large proportion of the ^{125}I-iodorotenone washes out of the myocardium much later. The results of the PESA analysis therefore suggest that ^{125}I-iodorotenone is preferentially retained in the rabbit myocardium and may have potential as an improved flow tracer. The high variance of coefficient c_1 for ^{99m}Tc-sestamibi cannot be explained at the present time but is likely due to a source of error in the experimental process.

The validity of the inferences drawn from the mean spectral coefficient vectors is dependent on the curvature of the likelihood function at each of the 25 solutions. Figure 13

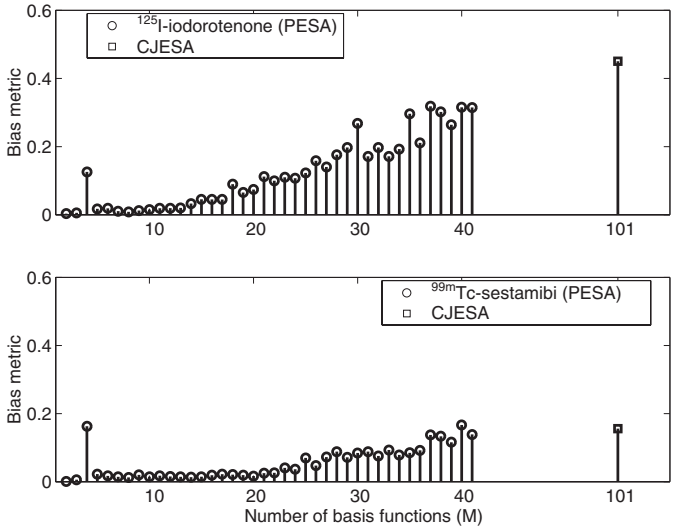


Figure 11. Bias metric B_M over $K = 100$ simulations. The value for the CJESA estimate appears at $M = 101$.

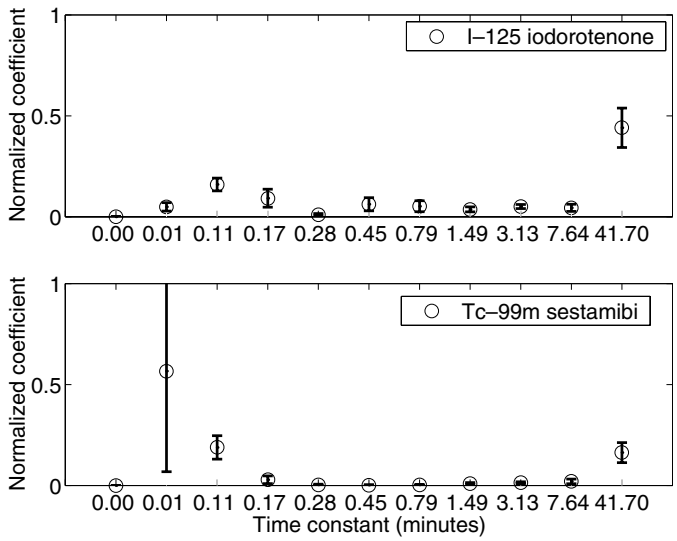


Figure 12. Normalized mean spectral coefficients for $M = 11$, shown $\pm$ the SEMs. These statistics are calculated from the $N = 25$ solutions that make up each mean.

shows the normalized mean vectors again, but with the error bars corresponding to the $\mathbf{s}_I = \mathbf{s}_{11}$. The coefficient c_1 of the ^{99m}Tc-sestamibi mean exhibits the 'loosest' bound. However, this uncertainty is only approximately 13% of the SEM for this coefficient. It consequently does not invalidate the inferences drawn above (especially since the difference in coefficient c_{11} is highly significant), but does reduce the statistical significance of the difference in this coefficient.

The execution times for PESA on a Pentium III 850 MHz processor appear in table 3.

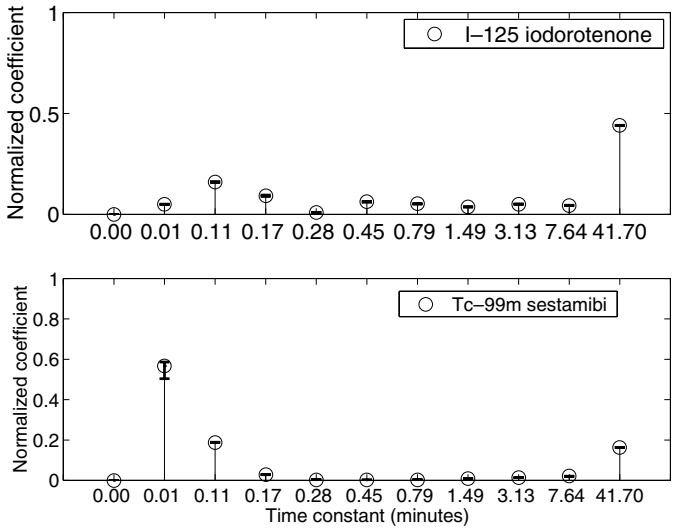


Figure 13. Normalized mean spectral coefficients for $M = 11$, shown $\pm \mathbf{s}_{11}$. The latter indicate the contributions to the uncertainty of these mean vectors of the intrinsic uncertainties at each of the solutions that make up the mean.

Table 3. Execution times for PESA for analysis of isolated rabbit heart experimental data (M = 1 to 20)

Algorithm phase	Value	Unit
Find M_* criteria	27	s
Optimize basis sets	438	min
Fit spectral models (50 datasets)	2	min
Simulate models ($K = 100$)	53	min

7. Conclusion

We have described several advantages of ESA over nonlinear parameter fitting methods for the modelling of the superposed responses of several first-order linear systems. The original fixed-basis ESA method proposed by Cunningham and Jones was found to yield parameter estimates that do not in general allow the comparison of experimental results using the spectral coefficients. This is because the presence of redundant parameters leads to fitting of the model to the noise within the measurement data. Solutions fluctuate considerably and introduce unacceptable bias into the estimates. We have presented PESA as one possible practical approach to this problem. The key idea of PESA is to approximate a redundant ESA basis with one of lower dimension by optimizing the choice of the k_2 values of each basis element. The basis dimension selection process then allows a compromise to be found between model approximation error and parameter bias. Where a suitable compromise cannot be achieved for a specific application, it is likely that the superposed first-order response model used by PESA is inappropriate for the data. Problems may also be encountered when the SNRs of two sets of data to be compared differ by a large amount, necessitating a different selection of M for each set.

Once a solution is obtained, we have shown how a constrained Cramér–Rao lower bound may be employed as a measure of the uncertainty in cases where the mean of several spectral estimates must be calculated. Ideally, the mean spectral vector should be constituted as a weighted average of the spectral coefficients of the N experiments. These weights might be based on the cCRLBs of the N solutions that contribute to the mean. However, owing to the fact that the cCRLB vectors typically differ in dimension, it is not clear how this weighting may be easily achieved. Appropriate methods are currently under investigation.

We have used the parameter bias as estimated via repeated simulation of the data as a criterion of solution stability. This has the disadvantage of being sensitive to errors in the noise model, and may be time consuming in some applications. In cases where this poses difficulties, a cross-validation procedure may be employed to determine if a model fitted to a partition of the data generalizes well and accurately models a non-overlapping partition of the dataset.

In application to simulated data, PESA produced spectral estimates exhibiting far lower bias than the Cunningham and Jones method, while achieving a low approximation error magnitude in comparison to the contributions of other sources of uncertainty in typical physical measurements. When applied to time-activity data obtained from radiotracer perfusion experiments on isolated rabbit hearts, PESA yielded parameter estimates that facilitated immediate comparison of the biokinetic profiles of the two perfused compounds in terms of tracer retention in the myocardium.

The most computationally intensive phases of the PESA algorithm are those involving the optimization of the basis elements and the simulation of the fitted models. The basis optimization time becomes inconveniently large for $M > 15$. However, since the optimized k_2 values may be easily reused through rescaling, this phase need not be carried out for every new application of PESA. Consequently, the applicability of PESA to large-scale problems is limited rather by the time-consuming simulation stage. When PESA is applied to large datasets, such as the TACs that make up high-resolution dynamic image sequences, it may be prudent to perform the bias estimation phase on a small random sample of the pixel time-activities.

Based on the evaluations we have performed, we believe the PESA basis should be considered for use in dynamic emission computed tomography (dECT) reconstruction algorithms that currently employ either convolved orthogonal basis functions (COBFs), oblique-rotated COBFs (ORCOBFs) or B-splines (Maltz 2000, 2001, Sitek *et al* 1999, Reutter *et al* 2000). While the use of the PESA basis is expected to marginally increase the basis dimension over that required for COBFs and ORCOBFs, it offers the potential advantage of yielding easily interpreted parametric images describing kinetic parameters directly from projection data. Improving the visual representation of dynamic reconstructions in this way may help dECT methods gain wider clinical acceptance.

Another medical imaging application for which PESA may be well suited is the reconstruction of relaxographic magnetic resonance images (Labadie *et al* 1994, Lee *et al* 1999). A relaxographic image represents the spatial distribution of spin relaxation decay times within a magnetic resonance (MR) signal. These decay times (corresponding to the time constants of real decaying exponentials) are recovered from the MR signal through application of inverse Laplace transform methods. In cases where several different exponential decays occur in one image voxel, PESA may be used to perform this inversion. The parametric PESA representation may be preferable to the non-parametric output of other methods (Lee *et al* 1999) in some applications.

PESA may also be preferable to the robust and widely employed methods developed by Provencher (Provencher 1976, 1982, Provencher and Dovi 1979) in applications where

(i) the results of several experiments are to be compared in terms of their representation under a single basis set,
(ii) the smoothness of the spectrum is unknown *a priori*,
(iii) deconvolution of the input function and exponential sum will introduce significant error.

The methods of Provencher should be used in preference to PESA in exponential sum fitting applications where parsimony and the attainment of high approximation accuracy are the principal objectives, rather than parsimony and the standardization of the basis employed.

Acknowledgments

The author would like to thank Robert Marshall, Ronald Huesman, Patricia Powers-Risius and Bryan Reutter for providing the isolated rabbit heart data and assisting in its analysis. Thanks are also due to Thomas Budinger, who motivated this work and assisted in the editing of this paper, to Jeffery Fessler and Jinyi Qi for valuable comments, as well as to Matthew Darmalingum, who produced the illustrations and Robert Smith, who proofread the manuscript.

This work was supported in part by the National Heart, Lung and Blood Institute of the US Department of Health and Human Services under grants HL-07367, R01-HL50663 and P01-HL25840, and in part by the Director, Office of Science, Office of Biological and Environmental Research, Medical Sciences Division of the US Department of Energy under contract DE-AC03-76SF00098.

References

Ang D D, Lund J and Stenger F 1989 Complex variable and regularization methods of inversion of the Laplace transform *Math. Comput*. **53** 589–608

Bellman R 1966 *Numerical Inversion of the Laplace Transform* (New York: Elsevier)

Braess D 1986 *Nonlinear Approximation Theory* (Berlin: Springer)

Cunningham V J and Jones T 1993 Spectral analysis of dynamic PET studies *J. Cereb. Blood Flow Metab.* **13** 15–23

Dong C W 1993 A regularization method for the numerical inversion of the Laplace transform *SIAM J. Numer. Anal*. **30** 759–73

Gardner D, Gardner J, Laush G and Meinke W 1959 Method for analysis of multicomponent exponential decay curves *J. Chem. Phys*. **31** 978–86

Gorman J D and Hero A O 1990 Lower bounds for parametric estimation with constraints *IEEE Trans. Inf. Theory* **26** 1285–1301

Labadie C, Lee J H, Vetek G and Springer C S Jr 1994 Relaxographic imaging *J. Magn. Reson*. B **105** 99–112

Lanczos C 1956 *Applied Analysis* (Englewood Cliffs, NJ: Prentice-Hall)

Lee J H, Labadie C, Springer C S Jr and Harbison G S 1993 Two-dimensional inverse Laplace transform NMR: altered relaxation times allow detection of exchange correlation *J. Am. Chem. Soc*. **115** 7761–4

Lee J H, Li X, Sammi M K and Springer C S Jr 1999 Using flow relaxography to elucidate flow relaxivity *J. Magn. Reson*. **136** 102–13

Maltz J S 2000 Direct recovery of regional tracer kinetics from temporally inconsistent dynamic ECT projections using dimension-reduced time-activity basis *Phys. Med. Biol*. **45** 3413–29

Maltz J S 2001 Optimal time-activity basis selection for exponential spectral analysis: Application to the solution of large dynamic emission tomographic reconstruction problems *IEEE Trans. Nucl. Sci*. **48** 1452–64

Marshall R C, Powers-Risius P, Reutter B W, Taylor S E, VanBroklin H F, Huesman R H and Budinger T 2001 Kinetic analysis of ^{125}I-iodorotenone as a deposited myocardial flow tracer: Comparison with ^{99m}Tc-sestamibi *J. Nucl. Med.* **42** 272–81

McWhirter J G and Pike E R 1978 On the numerical inversion of the Laplace transform and similar Fredholm integral equations of the first kind *J. Phys. A: Math. Gen*. **11** 1729–45

Miller K M and Guy J W T 1966 Numerical inversion of the Laplace transform by use of the Jacobi polynomial *SIAM J. Numer. Anal*. **3** 624–35

Osborne M R and Smyth G K 1991 A modified Prony algorithm for fitting functions defined by difference equations *SIAM J. Sci. Stat. Comput.* **12** 362–82

Prony R 1795 Essai expérimental et analytique: sur les lois de la dilatabilité de fluides élastiques et sur ce de la force expansive de la vapeur de l'eau et de la vapeur de l'alcool, à differentes températures *de L'Ecole Polytechnique* **1** 24–76

See endnote 1

Provencher S W 1976 An eigenfunction expansion method for the analysis of exponential decay curves *J. Chem. Phys.* **64** 2772–7

Provencher S W 1982 A constrained regularization method for inverting data represented by linear algebraic or integral equations *Comput. Phys. Commun.* **27** 213–27

Provencher S W and Dovi V G 1979 Direct analysis of continuous relaxation spectra *J. Biochem. Biophys. Methods* **1** 313–18

Ramm A G 1986 Inversion of the Laplace transform *Inverse Probl.* **2** 55–59

Reich J G 1981 On parameter redundancy in curve fitting of kinetic data *Kinetic Data Analysis: Design and Analysis of Enzyme and Pharmacokinetic Experiments* ed L Endrenyi (New York: Plenum) ch IV pp 39–60

Reutter B W, Gullberg G T and Huesman R H 2000 Direct least squares estimation of spatiotemporal distributions from dynamic cardiac SPECT projections using spatial segmentation and temporal B-splines *IEEE Trans. Med. Imaging* **19** 434–50

Schnedermann E 1994 The inverse Laplace transform as the ultimate tool for transverse mass spectra *Z. Phys.* C **64** 85–90

Sitek A, Di Bella E V R and Gullberg G T 1999 Factor analysis of dynamic structures in dynamic SPECT imaging using maximum entropy *IEEE Nuclear Science Symposium and Medical Imaging Conference Record* pp 2227–32

Steiglitz K 1977 On the simultaneous estimation of poles and zeros in speech analysis *IEEE Trans. Acoust. Speech and Signal Process.* **25** 229–34

Van Trees H L 1968 *Detection, Estimation, and Modulation Theory* vol I (New York: Wiley)

Varah J M 1983 Pitfalls in the numerical solution of linear ill-posed problems *SIAM J. Sci. and Stat. Comput.* **4** 164–76

Widder D V 1941 *The Laplace Transform* (Princeton, NJ: Princeton University Press)

Endnotes

(1) Author: Please check the year of publication in ‘Prony 1795’.